Efficient Video Breakup Detection and Verification

Martin Winter
JOANNEUM RESEARCH
Forschungsgesellschaft mbH
DIGITAL - Institute of
Information and
Communication Technologies
Steyrergasse 17, A-8010
Graz, Austria
martin.winter@joanneum.at

Peter Schallauer
JOANNEUM RESEARCH
Forschungsgesellschaft mbH
DIGITAL - Institute of
Information and
Communication Technologies
Steyrergasse 17, A-8010
Graz, Austria
peter.schallauer@joanneum.at

Albert Hofmann
JOANNEUM RESEARCH
Forschungsgesellschaft mbH
DIGITAL - Institute of
Information and
Communication Technologies
Steyrergasse 17, A-8010
Graz, Austria
albert.hofmann@joanneum.at

Hannes Fassold
JOANNEUM RESEARCH
Forschungsgesellschaft mbH
DIGITAL - Institute of
Information and
Communication Technologies,
Steyrergasse 17, A-8010
Graz, Austria
hannes.fassold@joanneum.at

ABSTRACT

Automatic quality assessment for audiovisual media is an important task for several steps of the media production, delivery and archiving processes. In this paper we focus on the semi-automatic quality inspection of videos and propose a novel algorithm for the detection of severe visual distortions, commonly termed as 'Video Breakup'. In order to enable the efficient human interaction with quality analysis results we present the extended 'Quality Summary Viewer' application, which enables the user to efficiently verify the detected events. Moreover it allows the user to quickly grasp the frequency and strengths of these visual impairments in the content and allows for its overall quality appraisal. Evaluation on a huge, challenging dataset shows, that our algorithm is able to detect up to 97.3% of the events annotated by domain experts. Depending on the application specific needs, we can reach a false detection rate of only 0.1-1.5 per minute.

Categories and Subject Descriptors

Computing Methodologies [**IMAGE PROCESSING AND COMPUTER VISION**]: Applications

General Terms

Algorithms, Verification

AIEMPro'10, October 29, 2010, Firenze, Italy.

Keywords

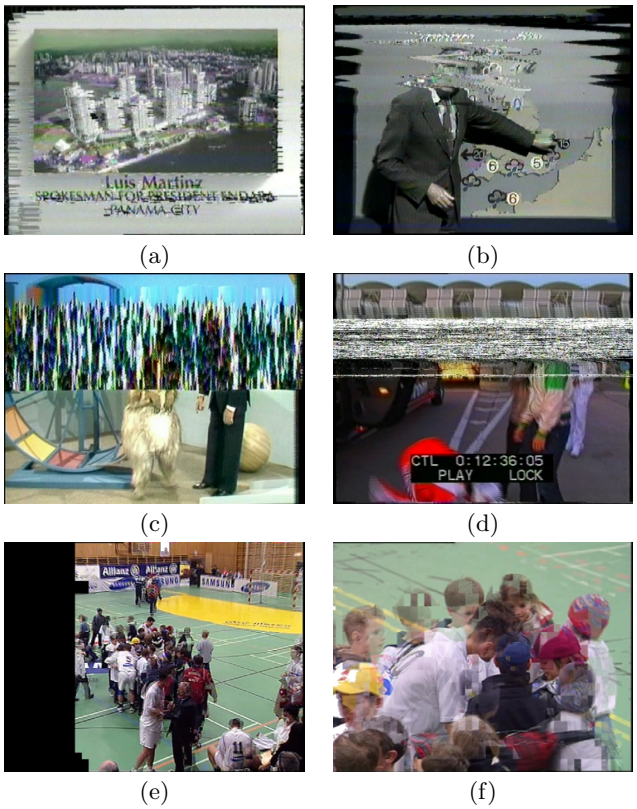
Preservation, broadcast, archives, video, cultural heritage, production, video breakup, visual impairment, summarization, quality assessment.

1. INTRODUCTION

Automatic quality analysis of audiovisual content is an important task for several steps of the media production, delivery and archiving process and affects multiple domains. For instance, content providers are checking post production content for correct encoding and conformance to the required quality and format standards before dispatching to the broadcasters or other service providers. Broadcasters on their part are checking audio and video quality material as part of the ingest process, after editing, after encoding and before play-out for terrestrial, satellite and cable broadcast or for delivery to internet and video-on-demand services.

Within the digital video and film preservation application domain the results of content based quality analysis aim at improving efficiency of various archive related processes. They check the content integrity during ingest process, perform 'best copy selection' when multiple copies of material are available and provide minimum quality service for archive accesses.

Although there is such a widely spread band of applications, it is noticeable, that currently mainly the technical properties of the video material are checked (*e.g.* stream compliance, GOP structure, playtime, aspect ratio, resolution or MXF compliance). Thus automatic content based quality control is currently only limited to a few number of simple measures as *e.g.* amount of blocking, detecting the existence of luma/chroma violation or rough estimation of noise levels. Other relevant content properties and impairments are nowadays usually checked by humans manually exploring the audiovisual content, which is a very labor intensive, tedious and expensive work. There has been only



(a) (b) (c) (d) (e) (f)

Figure 1: Some typical examples for video breakup impairments caused by analog (a-d) and digital (e,f) sources.

limited research regarding automatic video breakup detection for quality control performed so far. The most related paper to our approach is the one from Wang and Li [5] describing temporal smoothness constraints of consecutive video-frames' wavelet decomposition.

So in this work we focus on the detection of this very important type of content based visual defects, namely 'severe visual distortions', which are commonly also known as 'video breakups'. They can be roughly divided into two main groups with respect to the origin of the error introduced: Analogue errors are typically caused by tape transportation (*e.g.* wrong position of a track) or video signal transmission problems, while digital errors are often caused by (partially) broken video streams. In figure 1 one can see some typical examples for 'video breakup' we are focusing on.

The rest of the paper is organized as follows. In section 2 we have a short look on the algorithms' requirements for the detection of content based quality analysis, describe the details of our basic 'video breakup' detection algorithm and introduce a quality enhanced (motion compensated) version of the algorithm. In the following section 3 we present the results of our extensive evaluation on a challenging set of videos. Finally, in section 4 we present our visualization framework, which serves as a prototypical application for semi-automatic quality inspection of videos.

2. 'VIDEO BREAKUP DETECTION'

2.1 General requirements on the algorithm

In order to design algorithms for the purpose of content based video impairment analysis it is important to understand the basic application requirements for those algorithms. Thus in the following we define some requirements for impairment detection algorithms, tools and systems for software and file based environments.

At first, the algorithms should work semi-automatic or even fully automatic, because the full, manual inspection requires at least the video's duration of labor time and is therefore very cost intensive. Moreover, as human concentrativeness usually weakens over time, the quality of inspection decreases and 'objective' judgment of video quality is not guaranteed for the whole period of inspection.

Another important issue is the requirement regarding runtime. For the broadcast delivery services it is essential to do quality inspection 'on the fly', which requires at strict realtime performance of the algorithms. Nevertheless, most archive and broadcast related processes allow offline quality inspection. Thus the algorithms runtime for analyzing a single video is not that critical. Anyway, the amount of content in audiovisual archives and broadcast production processes requires a high overall throughput. This can be usually achieved either by efficient or parallelized algorithms, or by distributing the analysis of the videos to multiple cores, CPUs or machines.

Regarding the performance of the system it is obvious, that the detection rate should be as high as possible. But also the number of erroneous detections (wrong detections, 'false positives') is very important for the acceptance of the system, because too much false alarms for severe distortions annoy not only the operator, but limit also the time saving (and thus cost benefit) gained by applying the algorithm.

For seamless integration with software based application environments, quality analysis should also be implemented as much as possible in software. Extensibility and flexibility of a software based implementation is preferable over a hardware based solution. In this context it is worth to mention, that working on raw video data is also preferable in order to avoid any special video encoding (encoder specific properties) dependency.

Finally the algorithm should provide abstracted, compact information about the quality of the content(*e.g.* timecode of occurrence, defect class and its strength). Only the abstract information can be visualized in a compact way, which is a pre-requisite for efficient human inspection of analysis results.

2.2 Basic 'Video breakup' detection algorithm

The appearance of the 'video breakup' defects varies substantially and therefore it is difficult to identify common patterns for reliable, severe distortion detection (see figure 1 for some examples). Thus it is not feasible to try an explicit modeling of the defect itself. Therefore we invert the problem, model the normal 'video sequence behavior' and detect 'video breakups' by violation of its continuous motion constraints.

As mentioned in section 2.1, impairment detection algorithms are subject to several design criteria where especially the runtime constraints prohibit the design of a complex and therefore time consuming algorithm. Fortunately we

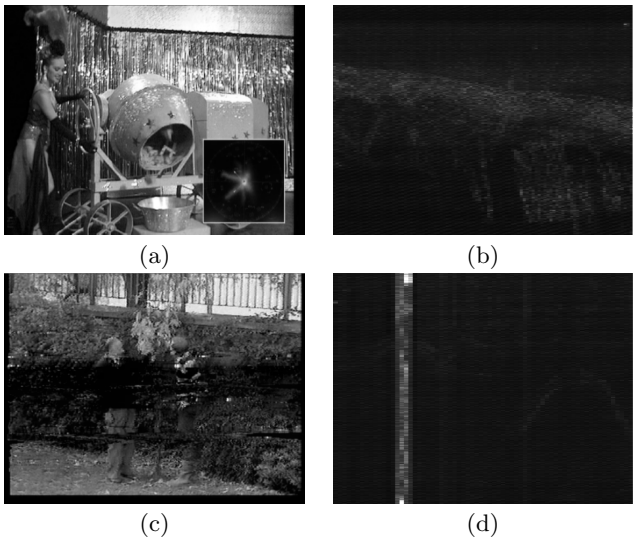


Figure 2: Some examples for 'stripe-difference-images' obtained (b,d) from a scene containing regular motion of a rotating, downwards moving concrete mixer(a) and a scene containing a typical video breakup impairment(c).

have observed, that - with the exception of fast motion sequences - even simple pixel differences between consecutive frames can distinguish between normal object motion within the content and abrupt content changes induced by 'video breakup'. This is, because 'video breakup' distortions significantly change their location and appearance from frame to frame which is usually not the case for regular moving objects. Thus we can design a very simple indicator for the presence of 'video breakup' events with surprisingly good performance results.

In particular let the actual image be I_n, then for each difference image $D_{n-1,n}$ between consecutive frames we cumulate the intensity difference values in each row of D and obtain a 'stripe-difference-image' representation as *e.g.* shown in Figure 2(b) and (d). While regular motion of the concrete mixer (downwards motion) causes only slight changes from column to column (b), 'video breakups' as *e.g.* shown in Figure 2(c) induce heavy disturbances. Hence, the H_6 distance measure from [4], calculated between consecutive columns of the 'stripe-difference-image', allows us to estimate the probability of video breakup presence for each frame of the video.

As we found, that this basic 'video breakup' measure proposed occasionally fails in scenes with heavy, suddenly appearing illumination changes (*e.g.* caused by flash lights or other strong illumination changes), we introduce a second measure based on the ratio R between vertical and horizontal edges on the difference image directly. In particular we estimate the averaged change of R within a certain time interval (5 frames) to indicate a 'video breakup' event. The basic assumption behind is, that the amount of horizontal and vertical edges changes equally for extensive illumination changes, while this is usually not the case for analogue 'video breakups'.

2.3 Motion compensated 'video breakup' detection

To overcome the limitations regarding fast motion sequences introducing false detections, we estimate the motion between two consecutive frames I_{n-1} and I_n and use the motion field obtained to warp the image I_{n-1} to the motion compensated image $I_{n_{mc}}$. Thus the motion compensated difference image $D_{n_{mc},n} = I_n - I_{n_{mc}}$ ideally contains only the not motion compensable part between two consecutive frames, which is usually significantly higher if 'video breakup' defects occur. In its original form, optical flow calculation has been a very time-consuming task and was therefore rarely used for applications with real-time requirements. Nevertheless, recent research takes advantage from the highly parallelized architecture of graphic processors (GPU) thus providing sufficient run-time capabilities for optical flow calculation even on full standard definition resolution videos [6].

3. EVALUATION

To evaluate the performance of our algorithm we use a compilation of 51 videos with various, challenging contents (*e.g.* scenes with extremely fast motion, heavy luminance changes, noise, water etc.), containing several hundreds of video impairments and an overall amount of 452 minutes content. For an objective judgment, the video breakups have been annotated by experts from a local video producing company. Besides the exact location within the video, for each video breakup a 'subjective' strength (6 levels) has been recorded, so that evaluations focusing on the strengths of the impairments are possible. As evaluation metrics we use the well known recall-measure R defined by $R = \frac{TP}{TP+FN}$, where TP denotes the number of 'correct detections' (true positives) and FN counts the number of 'missed detections' (false negatives) within a video. It estimates the fraction of impairments correctly detected overall. As contrary measure, estimating the precision of the detection we use the false positive rate FPR defined by $FPR = \frac{FP}{t}$, where FP is the number of erroneous detections normalized by the time interval t. Note, that we consider the FPR to be more appropriate for the typical scenario than the classical 'precision' measure $P = \frac{TP}{TP+FP}$, as it is not weighted by the number of true detections (TP) and directly reflects the displeasedness felt by an operator, when obtaining a significant amount of erroneous detections. In contrast to binary event evaluations, where time-dependency is not a factor (yes/no decisions), we have to take special care for 'impairment' events which typically range over a certain time interval. Thus we treat a video breakup event as correctly identified (TP), if there is at least one response of our algorithm within a time segment annotated in the ground-truth.[1] An erroneous detection is consequently reported, if there is no corresponding ground-truth segment annotated within the time segment of our algorithms positive response.[2]

Figure 3(a) shows the overall performance results of our

[1]Although calculating an temporal overlap between the annotated and detected events might be a feasible measure, we decided to use the proposed measure as this better reflects operator/user needs.

[2]In particular we allow for a tolerance of 4 frames when searching for a corresponding response, as in practice, especially for short video breakups, inspection of the defects' vicinity is anyway done by the operator.

algorithm for the basic (STD) and motion compensated version (ACC) of our Video Breakup algorithm. As one can see, the overall difference between the basic and motion compensated version characteristics is not that high, but especially for higher recall rates (above *e.g.* 95%) the difference in the 'precision' measure and thus the performance gain with respect to the erroneously detected video breakup events is evident (0.8 vs. 1.3 false positives min^{-1}). In order to allow an operator for an easy selection of the desired analysis systems' behavior, we have defined 4 preset working points denoted in figure 3(a) as triangles and explicitly listed in table 1. Hence for the basic version as well as for the motion compensated version of our algorithm it is possible to reach nearly 100% recall (R), but with an significant difference in false detection rate. Depending on the 'subjective' strength annotated, we can also have a look on the performance measures if we take into account only the 'strongest' video breakups. The performance results for that case are depicted in Figure 3(b) and it can be seen, that the results are significantly better, although the differences between basic version and motion compensated version of our algorithms are smaller.

Table 1: Preset working points and corresponding recall and precision values taking into account all video breakups annotated in the ground truth for the basic (STD) and motion compensated (ACC) version of our proposed algorithm. Settings: 1 = *high recall*, 2 = *default*, 3 = *low false positives*, 4 = *extremely low false positives*.

mode	working point	recall	FPs per min
STD	1	42.70	0.098
	2	83.33	0.583
	3	93.03	1.044
	4	97.24	2.202
ACC	1	53.85	0.106
	2	82.23	0.437
	3	94.23	0.776
	4	97.29	1.518

Regarding runtime we made some speed tests on an Intel Core2 Quad CPU (2.4GHz) with 3GB of RAM and a NVIDIA GeForce GTX 285 (240 unified shaders) on standard (SD, 720 × 576) and high definition (full HD, 1920 × 1080) video material. The obtained run-times are summarized in table 2. On a standard definition video the motion compensated version needs 4 times the duration of the video for processing, while the basic version processing is faster than realtime and thus the only one candidate for usage in realtime requirement (*e.g.* broadcasting) scenarios. For full HD videos the run-time requirements increases quadratically up to 16-17 times the video length for the accurate mode and 3 times video length for the basic video breakup algorithm.

In order to get an information about the generalization ability of our evaluation scheme, we made an additional test constantly increasing the number of videos taken into account for an evaluation run and monitored the performance characteristics. In doing so we have observed that no significant change in performance values is obtained exceeding a set of about 30 files. This indicates that our results are quite stable and the results might be extrapolated to other videos, too. We have also experimentally checked the false positive rate on another 8 hours videos containing almost no video breakup impairments. The false positive rate decreased about one third compared to the results on our evaluation database. This fosters the assumption that our evaluation database has practical significance.

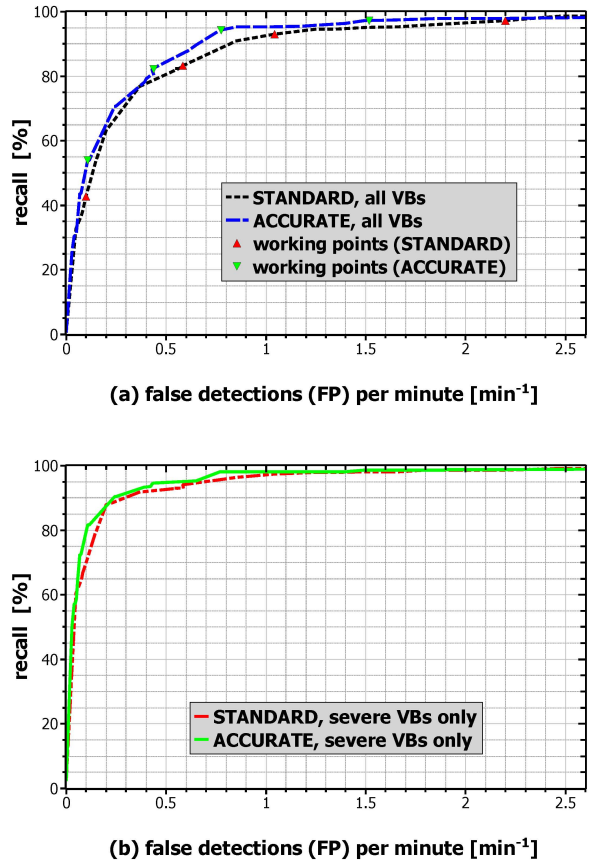


Figure 3: Performance results for the basic (STD) and motion compensated version (ACC) of our video breakup algorithm on a challenging dataset of 51 videos with about 452 min content overall. (a) shows the results taking into account all annotated video breakups in ground-truth, while (b) depicts the results only for the strongest (= subjective most significant) video breakups found.

4. VISUALIZATION AND EXCHANGE OF IMPAIRMENT DESCRIPTION

4.1 MPEG-7 description and classification

In order to facilitate inter-operability and exchange of impairment descriptions between different applications (*e.g.* automatic analysis tool and interactive visualization/verification application), a standardized way of defect description is needed. In previous work [3] we have already proposed a framework for the description of visual impairments based on MPEG-7 [2]. Based on the existing work in the audio part, we have defined a similar description framework for the visual do-

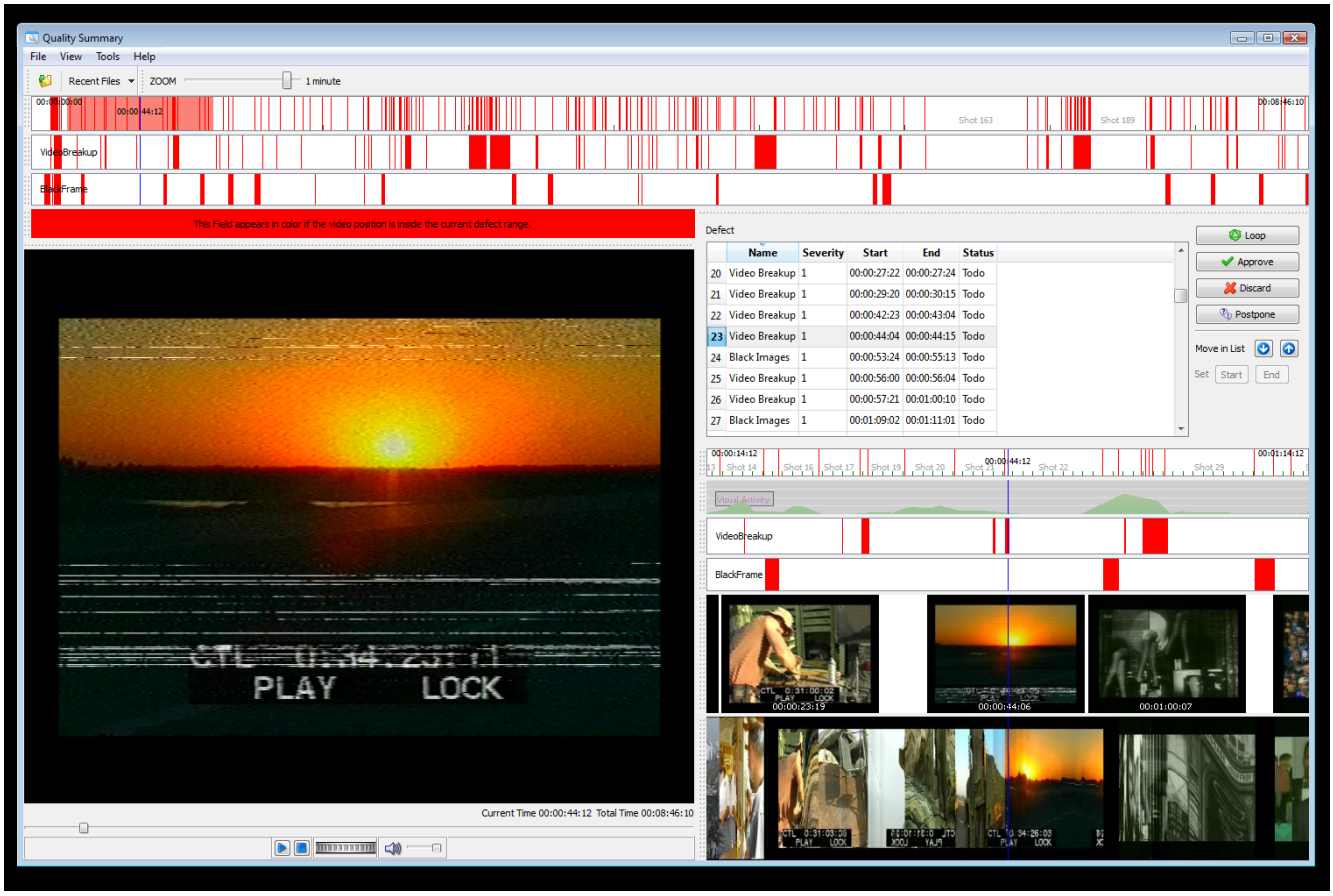


Figure 4: Screenshot of our Quality Summary Viewer. A list of defects detected lets the user quickly navigate from defect to defect. Time-line views show the shot structure of the material, selected representative key frames, stripe images and several impairment positions within the video.

Table 2: Summarized run-time results for on SD and full-HD video material. The 'real-time factor'(RT) is defined by the ratio of required analysis time and realtime duration of the video.

format/mode	RT	tot.(ms)	VB(ms)
SD, ACC	$0.25 \pm 9\%$	162.1	144.6
SD, STD	$1.47 \pm 1\%$	27.2	9.7
SD	$2.29 \pm 1\%$	17.5	-
HD, ACC	$0.06 \pm 14\%$	657.9	583.3
HD, STD	$0.31 \pm 2\%$	127.4	53.3
HD	$0.54 \pm 1\%$	74.3	-

main with even more capabilities for describing details of impairments. A list of defects (*e.g.* 'video breakups') and quality measures (*e.g.* noise/grain level) can be described for each segment, either using a generic descriptor identifying the impairment using a classification scheme or a specific descriptor for a certain impairment type. We have defined (and extended) a comprehensive impairment classification scheme that provides for hierarchical organization and multilingual description of impairments. The main organization criteria of the classification schemes are the visible and audible effects of impairments. The proposed MPEG-7 extension and classification schemes are available at [1].

4.2 Summarization and verification

The visualization and verification of defect analysis results must support the user in quickly getting an overview of the materials condition. For that purpose, we have extended our proposed 'Quality Summary Viewer' application [3] shown in Figure 4. This tool has been extended in order to allow visualization and direct navigation to specific 'video breakups' found within the video. The tool supports the user in efficiently navigating through the content by providing a time-line representation for a number of views. All views are synchronized with the video player and the temporal resolution can be changed so that the user can freely change the level of detail shown. The temporally condensed overview allows the user to quickly grasp the frequency and strengths of the 'video breakups' in the material. Besides the configurable visualization of various defect types, the time-line views show the shot structure of the material, selected representative key frames, stripe images created from the central columns of the images in the sequence and a number of configurable graphs visualizing the occurrence of several defects. A (sortable) list of impairments detected

lets the user quickly navigate from defect to defect. Hence, by investigation of the temporal neighborhood (loop playing around the defect one focuses on) a quick judgement about the 'video breakup' can be done. The user can either approve or discard the automatically detected events, without investigating the whole video content. Thus it is possible, to appraise the quality of the video much faster than by manual investigation of the whole video. Finally it is important to note, that the tool is very flexible, so that it is possible to easily integrate other impairment detection results or visualize local video quality measures (*e.g.* noise, visual activity, or other video statistics) in a common framework.

5. CONCLUSION

In this paper we have proposed a novel algorithm for the detection of severe visual distortions, commonly termed as 'Video Breakup', in videos. Furthermore, we presented a framework for visualization and easy judgment of the defects, so that it is possible to quickly get an overview about the materials condition. Depending on the domain of application it is possible to parameterize the algorithm meeting specific realtime requirements in standard definition (SD) resolution. Nevertheless, for a lower rate of erroneous detections we presented an improved version of the algorithm using optical flow for motion compensation on the GPU. An extensive evaluation on an expert's annotated, huge database shows, that we can detect up to 97.3% of the annotated video breakups. Depending on the application specific needs, we can reach a false detection rate of only 0.1-1.5 per minute. In the future we will especially improve the robustness of our algorithm with respect to ultra-fast motion scenes and multi-frame dissolves.

6. ACKNOWLEDGMENTS

The authors would like to thank Werner Haas, Georg Thallinger, Hermann Fürntratt and Marcin Rosner as well as several other colleagues at JOANNEUM RESEARCH and Martin Altmanninger and Paul Leitner from Media Services GmbH, who contributed valuable input to the work. Furthermore we give thanks to the Austrian Broadcasting Corporation (ORF), the British Broadcasting Corporation (BBC), RAI Italian Television and the Netherlands Institute for Sound and Vision (BnG) for providing requirement specifications as well as useful videos samples. This work has been funded partially under the 7th Framework Programme of the European Union within the IST project 'PrestoPRIME' (IST FP7 231161) and under the FIT-IT Programme of the Austrian Federal Ministry for Transport, Innovation and Technology within the project 'vdQA'.

7. REFERENCES

[1] Audiovisual Defect and Quality Description. http://mpeg7.joanneum.at.

[2] ISO/IEC. Multimedia content description interface. ISO/IEC 15938:2001.

[3] P. Schallauer, W. Bailer, R. Mörzinger, H. Fürntratt, and G. Thallinger. Automatic quality analysis for film and video restoration. In *IEEE ICIP*, San Antonio, USA, October 2007.

[4] D. Van der Weken, M. Nachtegael, and E. Kerre. Using similarity measures for histogram comparison. *Lecture Notes in Computer Science*, 2715:1–9, 2003.

[5] Z. Wang and Q. Li. Statistics of natural image sequences: temporal motion smoothness by local phase correlations. *Human Vision and Electronic Imaging XIV*, 7240:72400W, 2009.

[6] C. Zach, ThomasPock, and H. Bischof. A duality based approach for realtime tv-l1 optical flow. In *Proceedings of the 29th DAGM Symposium on Pattern Recognition*, pages 214–223, 2007.